A REVIEW ON AUTOMATIC SPEECH RECOGNITION SYSTEM

*SEERAT SHAKEEL WANI & **HENNA ARORA

*M.Tech. Research Scholar, Universal Group of Institutions Lalru,Mohali Punjab, India
**Assistant professor, Universal Group of Institutions Lalru,Mohali Punjab, India

ABSTRACT

Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore persons would also like to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices. At present, inputting the data in computer by speech and converting the data into another form for e.g. Text. with the help of automatic speech recognition system and its recognition by the computer is one of the developed scientific fields. As each language has its specific feature, the various speech recognition systems are investigated for the different languages. In this paper, we have classified the Automatic Speech Recognition System

KEYWORDS: ASR system Connected Words.

I. OVERVIEW OF AUTOMATIC SPEECH RECOGNITION SYSTEM (ASR)

The task of ASR is to take an acoustic waveform as an input and produce output as a string of words. Basically, the problem of speech recognition can be stated as follows. When given with acoustic observation $X = X_1, X_2 \ldots X_n$, the goal is to find out the corresponding word sequence $W = W_1, W_2 \ldots W_m$ that has the maximum posterior probability $P(W|X)$ expressed using Bayes theorem as shown in equation (1). The following figure 1 shows the overview of ASR system.[1]
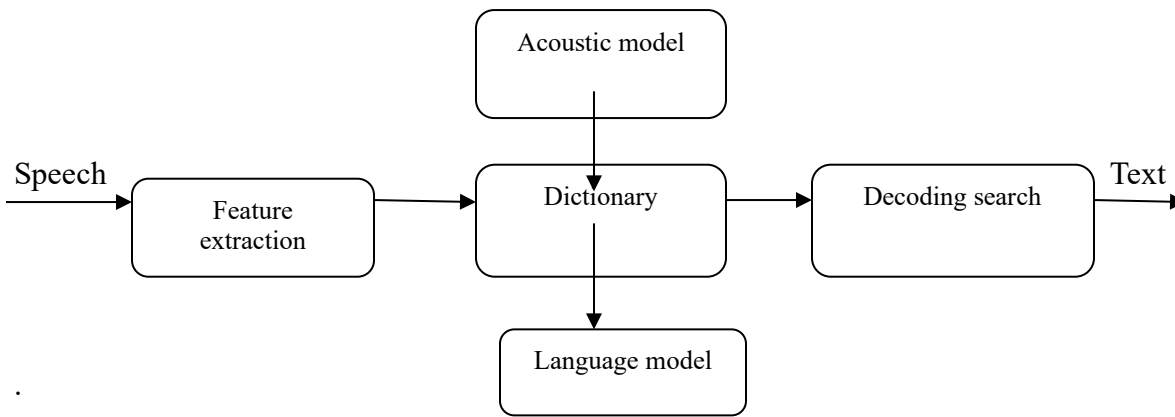
Figure 1 Overview of ASR systems

$$W = \text{argmax}_w \, P(W/X) = \text{arg max}_w \, P(W) \, P(X/W) \, / \, P(X) \tag{1.1}$$

Where P (W) is the probability of word W uttered and P(X|W) is the probability of acoustic observation of X when the word W is uttered In order to recognize speech, the system usually consists of two phases. They are called pre-processing and post-processing. Pre-processing involves feature extraction and the post-processing stage comprises of building a speech recognition engine.

II. DIFFICULTIES IN CONSTRUCTION OF ASR

Automatic speech recognition involves a number of disciplines such as physiology, acoustics, signal processing, pattern recognition, and linguistics. The difficulty of automatic speech recognition is coming from many aspects of these areas.[2]

Variability from speakers: A word may be uttered differently by the same speaker because of illness or emotion. It may be articulated differently depending on whether it is planned read speech or spontaneous conversation.

Variability from environments: This is because of background noise, reverberation, microphones, and transmission channels.

III. APPLICATION OF AUTOMATIC SPEECH RECOGNITION

Automatic Speech Recognition Systems are being used in many applications such as.
- Speech to text conversion using large vocabulary speech recognition systems.
- Message typing in mobile phones.
- Time and Attendance Systems.

- Access Control Systems.

- Telephone-Banking/Booking.

IV. CLASSIFICATION OF SPEECH RECOGNITION SYSTEMS

Speech recognition systems can be separated in several different classes by describing the type of speech utterance, type of speaker model, type of channel and the type of vocabulary that they have the ability to recognize.

V. TYPES OF SPEECH UTTERANCE

An utterance is the vocalization (speaking) of a word or words that represent a single meaning to the computer. Utterances can be a single word, a few words, a sentence, or even multiple sentences. The types of speech utterance are [3]

1) Isolated Words

Isolated word recognizers usually require each utterance to have quiet on both sides of the sample window. It doesn't mean that it accepts single words, but does require a single utterance at a time.

2) Connected Words   Connected word systems (or more correctly 'connected utterances') are similar to isolated words, but allow separate utterances to be 'run-together' with a minimal pause between them.

3) Continuous Speech

Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Basically, it's computer dictation.

4) Spontaneous Speech

This type of speech is natural and not rehearsed.

VI. TYPES OF SPEAKER MODEL

All speakers have their special voices, due to their unique physical body and personality. Speech recognition system is broadly classified into two main categories based on speaker models namely speaker dependent and speaker independent.[4]

1) Speaker dependent models

Speaker dependent systems are designed for a specific speaker. They are generally more accurate for the particular speaker, but much less accurate for other speakers. These systems are usually easier to develop, cheaper and more accurate, but not as flexible as speaker adaptive or speaker independent systems.

2) Speaker independent models

Speaker independent systems are designed for variety of speakers. It recognizes the speech patterns of a large group of people. This system is most difficult to develop, most expensive and offers less accuracy than speaker dependent systems. However, they are more flexible.

VII. TYPES OF VOCABULARY

The size of vocabulary of a speech recognition system affects the complexity, processing requirements and the accuracy of the system.[5] Some applications only require a few words (e.g. numbers only), others require very large dictionaries (e.g. dictation machines). In ASR systems the types of vocabularies can be classified as follows.

- Small vocabulary - tens of words
- Medium vocabulary - hundreds of words
- Large vocabulary - thousands of words
- Very-large vocabulary - tens of thousands of words
- Out-of-Vocabulary- Mapping a word from the vocabulary into the unknown word

Apart from the above characteristics, the environment variability, channel variability, speaking style, sex, age, speed of speech also makes the ASR system more complex. But the efficient ASR systems must cope with the variability in the signal. [6][7]

VIII. CONCLUSION

The major contribution of this work is a review   of the Automatic Speech Recognition (ASR) system. This work explains the application of ASR system and types of speaker model, types of vocabulary and also give the idea about the difficulties in construction of ASR system.

REFRENCES

[1]   B.S. Yalamanchili1, Anusha.K. K2, Santhi.K3, Sruthi.P4, SwapnaMadhavi.B5" Non Linear Classification for Emotion Detection on Telugu Corpus" B.S. Yalamanchili et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2443-2448

[2]   Brookes, M., Voicebox: Speech Processing Toolbox for Matlab [on line], Imperial College, London, available on the World Wide Web: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[3]   C. Chelba, T.J. Hazen, and M. Saraclar," Retrieval and Browsing of Spoken Content", IEEE Signal Processing Magazine 25 (3), May 2008

[4]   Daniel Jurafsky, and James H. Martin," Speech and language Processing", Pearson Education, 2000

[5]   Douglas O'Shaughnessy, "Interacting with Computer by Voice Automatic Speech Recognition and Synthesis", Proceeding of the IEEE, Vol.91, No.9, pp.1272-1305, Sept 2003

[6]   Frederick Jelinek, "The Dawn of Statistical ASR and MT, Computational Linguistics", Vol.35, No. 4, pp. 483-494, Dec 2009.

[7]   Guodong Guo and Stan Z. Li,"Content based audio classification and retrieval by SVMs", IEEE   trans. Neural Network, Vol.14, pp.209-215, Jan 2003.